University of Bath

MRES

A hybrid method for cellular invasion with realistic cell cycle distribution times

Young, Josh

Award date:
2019

Awarding institution:
University of Bath

Link to publication

Alternative formats
If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights
CC BY

Copyright of this thesis rests with the author. Access is subject to the above licence, if given. If no licence is specified above, original content in this thesis is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC-ND 4.0) Licence (https://creativecommons.org/licenses/by-nc-nd/4.0/). Any third-party copyright material present remains the property of its respective owner(s) and is licensed under its existing terms.

Take down policy
If you consider content within Bath's Research Portal to be in breach of UK law, please contact: openaccess@bath.ac.uk with the details. Your claim will be investigated and, where appropriate, the item will be removed from public view as soon as possible.

Download date: 05. May. 2023

A hybrid method for cellular invasion with realistic cell cycle distribution times

submitted by

Josh Young

for the degree of Master of Research

of the

University of Bath

Department of Mathematical Sciences

October 2019

Contents

Chapter 1

Introduction and literature review

1.1 Context

Biological systems exhibit a tremendously wide variety of behaviours at many different spatial scales. While in theory, the behaviour of a system at any scale can be viewed as emergent from the behaviour of its smallest components, deriving these scale relationships is analytically intractable except in only very carefully constructed examples. Numerical methods based purely on the microscopic behaviour of a system quickly become cost-prohibitive as the respective number of agents is increased. To counter this, computational feasibility may be attained through the use of scale-dependent modelling regimes; however, this inherently incurs a loss of information when coarser representations are applied. This report is concerned with hybrid methods, which combine multiple regimes to balance the advantages and disadvantages of each.

Many biological systems can be described as reaction-diffusion systems. These spatio-temporal models represent the evolution of particle systems via the incorporation of two distinct processes; in particular, inter-agent interactions, or reactions, and random movement, or diffusion. Such models are capable of displaying a multitude of behaviours, including wave-like behaviour and, in particular, the formation of travelling waves [12, 32]; as well as the emergence of self-organising spatial patterns [37], the formation of which are of great interest for explaining the structural heterogeneity present in most multicellular life. The broad applicability and generalisability of reaction-diffusion systems has motivated the development of many different modelling representations across the molecular, microscopic, mesoscopic, and macroscopic scales.

The finest modelling scale is that of molecular dynamics, in which the movement

of agents is simulated via consideration of their interactions with the media in which they are immersed. Coarse molecular representations treat agents and molecules as rigid spheres, calculating inter-molecular collisions according to standard Newtonian mechanics; that is, agents are allowed to collide and react in a manner which obeys conservation of momentum. In this way, these models do not consider diffusion to be a stochastic process, as in the coarser regimes described below, but rather a fully deterministic system of interacting agents. Inevitably, such models require the storage of the positions and momenta of all constituent agents, precluding their application to large systems.

The next coarsest ('microscopic') scale, consists primarily of agent-based models. Within this regime, the positions of all individual agents in the system are tracked; the diffusion of these agents can then be realised according to a Brownian motion or, in the case of biased or more complicated diffusive processes, a more general stochastic differential equation. A wide variety of methods exist for the modelling of inter-agent reactions, including the Smoluchowski coagulation equation [34], Green's function reaction dynamics [33], and the λ-ρ method [10]. For many systems, these methods provide the most accurate representation; however, this accuracy comes at a significant and potentially prohibitive computational cost. Keeping track of the location and movement of each individual agent quickly becomes very expensive for larger systems; modelling the diffusion of a system containing N agents requires the realisation of dN Gaussian random variables at each time step, where d is the dimension of the system. Further, the simulation of second- or higher-order reaction kinetics requires the storage of a, potentially very large, pairwise distance matrix in order to determine whether agents are close enough to react. This matrix must be recalculated at each time step, requiring $\mathcal{O}(N^2)$ operations for a system of N agents.

Mesoscopic methods, one of the two regimes of interest in this report, are more coarse than microscopic methods, sacrificing knowledge of individual agent locations in exchange for computational efficiency. Mesoscopic methodologies divide the spatial domain of interest into fixed-size compartments, and track the number of agents occupying each individual compartment; in particular, agent positions are known only up to their compartment, in which they are assumed to be well-mixed and uniformly distributed. Agents are allowed to interact at some prescribed rate within their compartment according to predetermined reaction pathways, typically manifesting as the addition or removal of agents from the compartment. In some cases, however, reaction products may be allowed to appear in neighbouring compartments [20]. Diffusion between compartments is treated as a reaction in which agents are removed from their initial compartment and placed into a neighbouring compartment; the corresponding

diffusion rate can be derived from the desired linear (as in Fick's law [11]) or nonlinear macroscopic diffusion law.

Compartment-based models can be simulated according to either a time-driven or an event-driven algorithm. Time-driven algorithms employ a fixed time step, chosen to be sufficiently small so as to preclude the occurrence of multiple reactions in any single time step. Both the occurrence or non-occurrence of a reaction, as well as the specific type of reaction which occurs, are determined by the generation of two uniformly distributed random numbers. Event-driven algorithms, on the other hand, simulate the elapsed time between reactions; the most widely used algorithm of this form is the Gillespie direct method [16]. All possible reactions have an associated exponentially distributed waiting time, representing the time until the reaction occurs; consequently, the time until the occurrence of any reaction is simply the minimum of these waiting times, which is itself exponentially distributed. A realisation of this time can then be drawn, with the associated reaction being chosen at random according to the relative individual reaction rates. While this method provides an exact simulation of the corresponding master equation, it can be expensive for large numbers of agents or compartments. In answer to this, several approximate stochastic simulation algorithms have been developed, such as the next reaction method [14], the next sub-volume method [8], and the R-leaping method [3].

Finally, macroscopic methods provide the coarsest applicable modelling framework, and shall be the second regime of interest in this report. These methods most frequently take the form of a partial differential equation such as the reaction-diffusion equation, given by

$$\frac{\partial \boldsymbol{u}}{\partial t} = \boldsymbol{D}\nabla^2 \boldsymbol{u} + R(\boldsymbol{u}), \tag{1.1}$$

where $\boldsymbol{u}$ represents agent concentrations, $\boldsymbol{D}$ is a matrix of diffusion and cross-diffusion coefficients, and R is a function representing the local reactions. Due to the extensive study of partial differential equations over the past several decades, there exist many highly-efficient methods for both simulation and solution, such as finite difference, finite element, and finite volume methods. The computational efficiency of these methods, however, comes at the cost of accuracy. In systems with regions of low agent density, PDE models are often unable to fully capture the nonlinear dynamics of the system.

The cell-division cycle is one of the most fundamental processes in biology, typically defined to be the period of time between a cell's formation through the division of some mother cell, and the division of the cell itself into two daughter cells. The vast majority of eukaryotes undergo some form of this cycle in order to replicate; prokaryotic cells, on the other hand, go through a comparatively simple process known as binary fission [1].

There are notable exceptions however, such as neuronal cells, which cease replication after a certain number of iterations, becoming permanently held in a particular phase of the cycle.

While the exact mechanisms behind division can vary substantially between cell types, the basic sequence of events remains essentially the same and can be broadly divided into four distinct phases: $\mathbf{G_1}$, $\mathbf{S}$, $\mathbf{G_2}$, and $\mathbf{M}$. The length of each phase is determined by a network of regulatory proteins which respond to internal and environmental cues, and exhibits considerable variation not only between different types of cell, but also between cells which are morphologically identical [1]. Measuring phase lengths is, moreover, a nontrivial experimental problem, as it is often not possible to observe all transitions between phases, or to identify which phase a cell is in at any given time; this further complicates the determination of cell cycle distributions [35].

1.2 Hybrid modelling

This section presents an overview of several hybrid methods, which couple the mesoscopic and macroscopic spatial regimes. Such methods divide the domain of interest into two or more regions, upon which different models are employed, in a manner which exploits their complementary strengths and minimises their weaknesses. All presented methods are described in the one-dimensional case; however, extension to multiple dimensions is, in most cases, trivial. Finally, we consider some existing methods applied in the study of the multi-stage model of the cell cycle.

We first introduce the pseudo-compartment method (PCM) of Yates and Flegg [38], which provides the basis for this report; a schematic diagram for this method is provided in Figure 1-1. The PCM works by dividing the domain of interest into two regions, separated by an interface. In one region, a PDE-based method is used, whilst a compartment-based method is employed in the other; these regions are labelled Ω_P and Ω_C, respectively. On Ω_P, we solve the PDE stated in Equation 1.1, representing the agent concentration at each point in the domain. This region is discretised into a regular grid with uniform spacing Δx, on which the PDE is solved using a finite-difference method. Specific details of the finite-difference scheme used for the solution of PDEs of this type will be given in Chapter 2.

The region Ω_C is uniformly divided into compartments of length h, where h is chosen such that $\Delta x << h$. The method keeps track of the number of agents in each compartment, but not the positions of these agents; we are therefore forced impose a uniform distribution and well-mixedness of the agents within each compartment. In each compartment, agents are allowed to either react according to some predetermined

reaction channels, or perform diffusive jumps into neighbouring compartments. The compartment-based method evolves according to the Gillespie direct method, a detailed discussion of which is presented in Chapter 2.

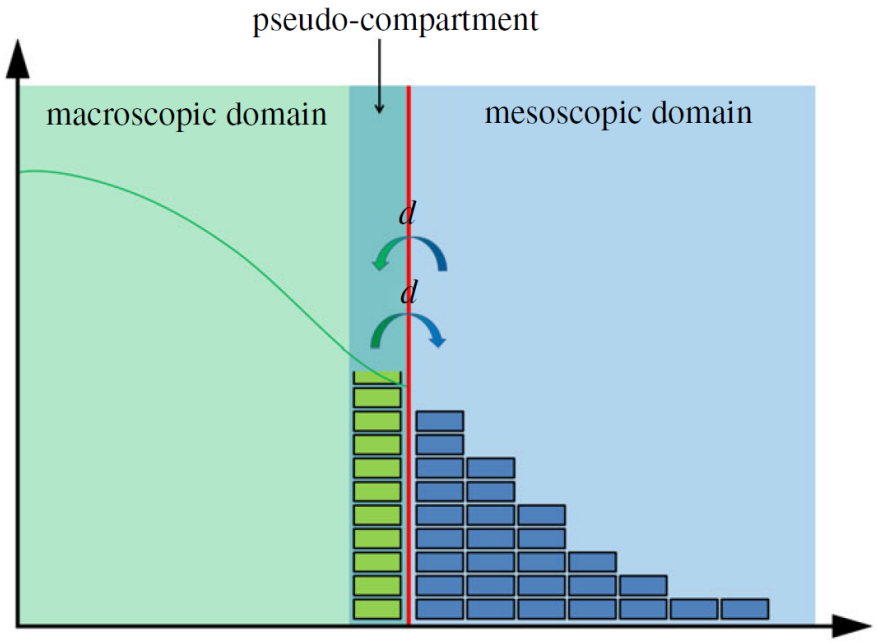

Figure 1-1: Schematic diagram for the PCM. The green line represents the solution of the PDE, while the blue boxes represent agents within each compartment. The red line denotes the interface between the two sub-domains. The green boxes residing in the pseudo-compartment represent the number of pseudo-agents within the pseudo-compartment, calculated by direct integration of the solution over that region. The arrows in the centre represent the movement of pseudo-agents over the interface between the pseudo-compartment and the first compartment of the mesoscopic domain. (Figure reproduced from [28].)

To facilitate the movement of agents between the two regimes, a pseudo-compartment is constructed, lying within the PDE-based region. This is a small region, of the same size as the other compartments, whose agent count is determined by direct integration of the solution of the PDE in said region. Agents are permitted to diffuse across the interface between the pseudo-compartment and Ω_C according to the Gillespie algorithm; this is achieved by either adding or removing a single agent's worth of mass from the solution of the PDE, uniformly across the pseudo-compartment, and removing or adding a agent to the first compartment accordingly. Other reactions, as well as diffusive events into Ω_P are not treated according to the compartment-based regime; rather, they are performed implicitly through solving the PDE. Chapters 3 and 4 will provide a much more comprehensive treatment of the PCM, wherein we will replicate

a travelling wave solution to the FKPP equation [12, 32] using this method, as well as an extension to the simulation of a multi-stage model of the cell cycle.

The PCM makes use of a static interface; that is, the interface and, in the case of the PCM, the pseudo-compartment, are stationary in space. While this simplifies the implementation of the method, it can limit the benefits of hybridisation when the agent mass in Ω_P and the pseudo-compartment is low. When agent numbers are low, the PDE-based method may fail to capture the full nonlinear dynamics of the system, due to the presence of bi-molecular or higher-order reactions in the underlying stochastic system. Further, in the case where Ω_C contains a large number of agents, the corresponding compartment-based method will be comparatively more expensive. Similar problems manifest in the hybrid coupling of alternative modelling regimes, such as microscopic-to-mesoscopic methods. To combat this, several hybrid methods have been developed which make use of adaptive interfaces; such interfaces may change either their position in space, or their size, according to the local distribution of agent mass.

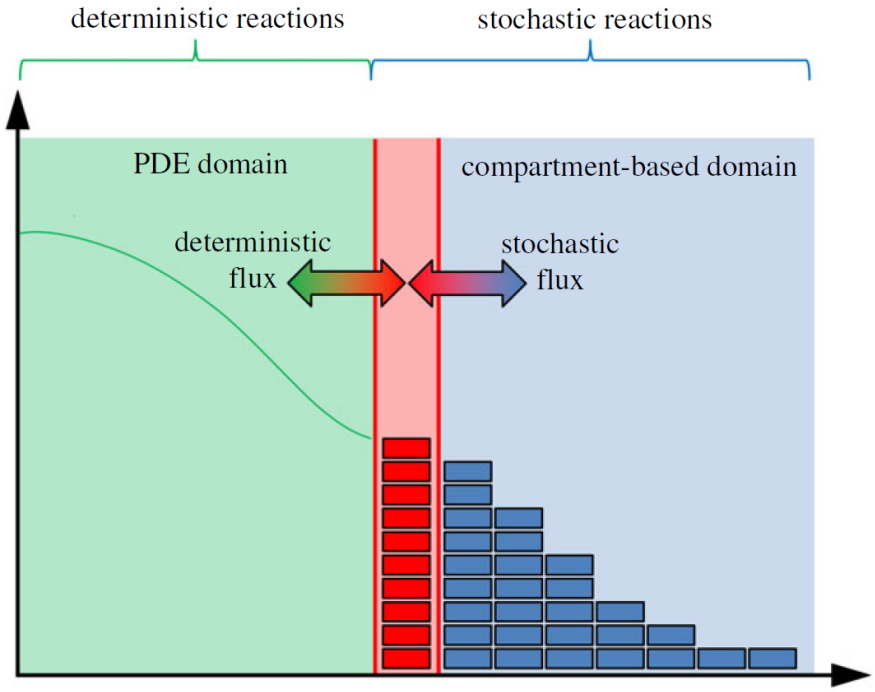

Figure 1-2: Schematic diagram for the method of Spill *et al.* [29]. The green line and blue boxes are defined in the same way as in Figure 1-1, while the red boxes denote an extra compartment between the PDE- and compartment-based regimes. The coloured double-headed arrows denote the manner in which fluxes are calculated between the two regimes. (Figure reproduced from [28].)

Spill *et al.* [29] incorporate an adaptive interface in their proposed method, which is otherwise conceptually very similar to the PCM; a schematic can be found in Figure 1-2. Rather than making use of a pseudo-compartment that is, in some sense, both a mesoscopic and a macroscopic region, their interface takes the form of a regular compartment. Further, the macroscopic region is divided into compartments, resulting in a system of mean-field ODEs representing the number of agents in each macroscopic compartment, as opposed to a PDE for the entire region.

Agents are permitted to jump into and out of the mesoscopic regime according to the stochastic simulation algorithm, whereas the macroscopic regime is coupled to the compartment according to a deterministic flux governed by the mean-field ODEs. The macro-compartments are labelled $C_{-n}, ..., C_{-1}$, the meso-compartments are labelled $C_1, ...C_m$, and the interface compartment is labelled C_0. An agent can jump from the first meso-compartment C_1 into the interface compartment C_0 at a rate proportional to the number of agents in C_1; likewise, an agent can jump from C_0 to C_1 at a rate proportional to the number of agents in C_0. Agents jumping between C_0 and C_{-1} are modelled deterministically. Letting $n(C_k, t)$ denote the number of agents in compartment k at time t, the flux between compartments C_0 and C_{-1} can be written

$$n(C_0, t+\tau) = n(C_0, t) + \tau \frac{D}{h^2}(n(C_0, t) - n(C_{-1}, t)), \tag{1.2}$$

where t is the current time, τ is the Gillespie time step, D is the diffusion coefficient, and h is the compartment width. While this appears to be a Neumann boundary condition at the interface, since $n(0, t)$ can change stochastically, this condition instead manifests as a stochastic source term for the mean-field ODEs at the interface.

All reactions within this compartment are completed according to the stochastic simulation algorithm, and reactions elsewhere are governed by their respective regime. A threshold number of agents is used to determine the location of, and adaptively move, this interface. The natural consequence of the imposition of this threshold is the occurrence of multiple interface regions, which the method permits. If a particular sub-domain of the mesoscopic region exceeds this threshold, for example, said sub-domain can be converted to a region governed by the PDE. Similarly, a sub-domain in the macroscopic region may spontaneously convert to a mesoscopic sub-domain. The ability for the domain to adapt its modelling regime according to local agent density incurs certain issues, however; in particular, there may be regions in which the agent density fluctuates quickly around the threshold value. While not an immediate problem, this could potentially give rise to numerous small, disconnected regions of a particular regime. This problem is circumvented by imposing a minimum permitted

size on sub-domains, preventing arbitrarily small regions from changing their modelling regime.

Further problems arise in the adaptive interface framework when regions of the modelling domain fluctuate about the threshold value. This causes the region in question to rapidly switch between modelling regimes, requiring a potentially significant amount of computational overhead. Further, a loss in accuracy may result. For example, in the event of a macroscopic region with a particularly high gradient switching to the mesoscopic regime, information about this gradient will be lost, introducing spurious smoothing of the solution. Robinson *et al.* [25] tackle this problem through the incorporation of two threshold values in their proposed hybrid method, which couples a macroscopic regime to a microscopic regime (we note, however, that their technique can be easily adapted to any coupling). In this methodology, if a microscopic region exceeds the upper threshold it switches to the macroscopic regime. Likewise, a macroscopic region which falls below the lower threshold will switch to the microscopic regime. Regions which lie between the two thresholds do not change their regime.

1.3 The cell-division cycle

This report considers the cell-division cycle, along with diffusion, as a mechanism for cellular invasion. A considerable research effort has been devoted to the study of the effects of cell motility and proliferation on invasion speed; in particular, deriving the relationship between motility and proliferation parameters and the invasion speed c. This began with the seminal work of Fisher [12] and, separately, Piscounov, Kolmogorov, and Petrovskii [32], who independently proposed the now ubiquitous FKPP equation,

$$\frac{\partial u}{\partial t} - D\frac{\partial^2 u}{\partial x^2} = ru(1-u), \tag{1.3}$$

with boundary conditions

$$\begin{aligned} u(-\infty, t) &= 1, \\ u(\infty, t) &= 0, \end{aligned} \tag{1.4}$$

where D is the diffusion rate, and r is the logistic growth rate. The FKPP equation was the first partial differential equation which was shown to exhibit the travelling wave behaviour expected from an invasive population.

Typical contemporary approaches for modelling invasion include deterministic models such as partial differential equations [36], and stochastic agent-based models (ABMs) [2, 7]. Recent work has extended these models to include cellular populations consisting

of multiple species or stages [9, 13]. It is well established that when agent motility is modelled according to standard Fickian diffusion [11], with diffusion coefficient D and assumed agent proliferation rate λ, the speed of invasion is proportional to $\sqrt{D\lambda}$ [12].

One key assumption of many representations of cellular invasion is that cells proliferate according to an exponentially distributed waiting time. Recent experimental results, however, indicate that this assumption is often inappropriate, and that cell-cycle time distributions (CCTDs) typically deviate significantly from an exponential distribution [4, 18, 39]. Instead, it has been proposed that CTTDs can be described with a considerably higher degree of accuracy through the use of hypoexponential distributions [4, 18, 39].

Multi-stage models (MSMs) are a class of model which seek to better describe CCTDs, which have garnered much attention in recent literature [39, 4]. MSMs divide the cell cycle into a sequence of N stages. Agents can proceed through the cycle from stage i to stage $i+1$ according to an exponentially distributed waiting time with rate λ_i; agents in the final stage N divide at rate λ_N, splitting into two daughter agents in stage 1, respectively. Analytic and numerical treatments of MSMs can be simplified through exploitation of the Markov property of the exponential distribution, resulting in hypoexponentially distributed CCTDs; two mathematical features which make MSMs appealing.

An important point to note is that the stages used in an MSM do not necessarily correspond to the biological phases of the cell cycle; rather, they are a mathematical construction which can be used to fit actual CCTDs. While MSMs have proven to be effective models in the case of well-mixed populations, little research has been conducted into the case of invasive populations, and the particular relationship between the parameters of MSMs and the resultant invasion speeds.

Recently, Gavagnin *et al.* [13] investigated the relationship between a multi-stage representation of the CCTD and invasion speed. In their paper, the authors constructed an ABM on a two-dimensional lattice, designed to represent cellular invasion. The proliferation rate of each agent was represented as an N-stage MSM that can be simulated using the Gillepsie method (or an alternative stochastic simulation algorithm). Through consideration of the average column densities and application of the mean-field approximation, they constructed a system of N reaction-diffusion-like PDEs. The authors then derived an exact analytic expression for the invasion speed in the case of identical inter-stage transition rates, corresponding to a CCTD with an Erlang distribution.

1.4 Summary

The overall purpose of this thesis formulation report (TFR) is to lay the foundations for building a hybrid method that couples the two modelling regimes used by Gavagnin *et al.* [13], exploiting the complementary advantages of both regimes. The ultimate aim is to construct a method exhibiting improved computational efficiency as compared to the ABM, whilst maintaining a similar level of accuracy. The TFR will also outline the motivation behind, and future research directions of, the PhD.

The structure of the remainder of this report is as follows. In Chapter 2, we present a detailed discussion of both the pseudo-compartment method of Yates *et al.* [38], and the multi-stage representation of CCTDs and invasion speeds of Gavagnin *et al.* [13]. For completeness, we additionally describe some of the mathematical background for these methods; in particular, we provide a representative example demonstrating the use of compartment-based modelling, and demonstrate how the mean-field assumption can be used to derive a corresponding partial differential equation. Further, we discuss techniques for simulating compartment-based models, such as the Gillespie direct method. The form and derivation of the general reaction-diffusion equation is explained, and the Crank-Nicolson finite difference method for solving general partial differential equations is presented.

In Chapter 3, we describe a multi-stage hybrid method for simulating cellular invasion with realistic cell cycle distribution times. The performance of this method is assessed for a simple test case, highlighting and interpreting the problems which arise. The report will conclude with Chapter 4, in which we discuss future research directions for the PhD.

Chapter 2

Methodology

This chapter explores in greater detail a number of concepts mentioned in the introduction. Specifically, we will first discuss compartment-based (mesoscopic) modelling and PDE-based (macroscopic) modelling, following which we will be sufficiently equipped to give a full explanation of the PCM. We will also replicate some of the results from the paper of Yates and Flegg [38], and demonstrate the capability of the method in the representation of more complicated dynamics. Finally, we introduce the mesoscopic and macroscopic methods of Gavagnin *et al.* [13] for the simulation of the cell-division cycle with diffusion, and realistic (non-exponential) cell cycle distribution times.

2.1 Mesoscopic modelling

A Markov chain approach is employed in the simulation of reaction-diffusion systems at the mesoscopic scale. The domain of interest is divided into a number of compartments, between which agents are permitted to jump and within which they can interact. This method keeps track of the number of agents inside each compartment, but does not consider individual agent locations within said compartment. With each possible diffusive jump and reaction channel we associate a propensity function, which represents the probability of said event occurring within an infinitesimal time step.

2.1.1 Compartment-based methods

As an illustrative example, consider dividing $\Omega = [0, 1]$ into K compartments of length $h = 1/K$. We denote by A_i the number of agents in compartment i, corresponding to the interval $C_i = [(i-1)h, ih]$, $i = 1, ..., K$. Each agent is permitted to jump out of its present compartment into a neighbouring compartment at some rate d. This can be

written as the following reaction system,

$$A_1 \underset{d}{\overset{d}{\rightleftharpoons}} A_2 \underset{d}{\overset{d}{\rightleftharpoons}} ... \underset{d}{\overset{d}{\rightleftharpoons}} A_K. \tag{2.1}$$

Further, each agent is allowed to either divide into two identical agents in the same compartment, or collide with another agent, destroying one of them. This can be written as the following reversible dimerisation reaction,

$$A_i + A_i \underset{k_2}{\overset{k_1}{\rightleftharpoons}} A_i, \tag{2.2}$$

for rates k_1, k_2 and $i = 1, ..., K$.

After enumerating all reaction channels, we can associate with the channel j a propensity function $a_j(t)$, defined such that $a_j(t)\mathrm{d}t$ is the probability of the reaction channel firing in the infinitesimal time step $[t, t+\mathrm{d}t)$. For reactions with a single reactant A_i, i.e. $A_i \xrightarrow{k} X$, where X is some arbitrary product, the corresponding propensity function is simply $kA_i(t)$, where $A_i(t)$ is the number of agents in compartment i at time t. For second-order reactions, such as $A_i + A_i \xrightarrow{k} X$, the propensity function is $\frac{k}{\nu}A_i(t)(A_i(t) - 1)$, where ν is the 'volume' of the compartment. In this example, where the compartments are one-dimensional, we simply have $\nu = h$.

We are now sufficiently equipped to determine the mean of the number of agents inside each compartment. For brevity, define the operators $R_i, L_i : \mathbb{N}^K \to \mathbb{N}^K$ as follows

$$\begin{aligned} R_i &: (n_1, ..., n_i, n_{i+1}, ..., n_K) \mapsto (n_1, ..., n_i + 1, n_{i+1} - 1, ..., n_K), \\ L_i &: (n_1, ..., n_{i-1}, n_i, ..., n_K) \mapsto (n_1, ..., n_{i-1} - 1, n_i + 1, ..., n_K). \end{aligned} \tag{2.3}$$

We then write down the chemical master equation (CME) [23]; this gives the time evolution of the probability that the system is in state $\boldsymbol{n} = (A_1 = n_1, ..., A_K = n_K)$ at time t. The CME for this system is as follows

$$\begin{aligned} \frac{\partial p(\boldsymbol{n})}{\partial t} = {} & d \sum_{i=1}^{K-1} \{(n_i + 1)p(R_i\boldsymbol{n}) - n_i p(\boldsymbol{n})\} + d \sum_{i=2}^{K} \{(n_i + 1)p(L_i\boldsymbol{n}) - n_i p(\boldsymbol{n})\} \\ & + \frac{k_1}{h} \sum_{i=1}^{K} \{n_i(n_i + 1)p(n_1, ..., n_i + 1, ..., n_K) - n_i(n_i - 1)p(\boldsymbol{n})\} \\ & + k_2 \sum_{i=1}^{K} \{(n_i - 1)p(n_1, ..., n_i - 1, ..., n_K) - n_i p(\boldsymbol{n})\}. \end{aligned} \tag{2.4}$$

The full derivation of the CME for a given system is somewhat laborious, and an exhaustive treatment can be found in [17]; here, it suffices to say the following. The

first and second summations represent the internal diffusive mechanisms through which compartments may arrive at the state $\boldsymbol{n}$ from a different state. The first way this could occur is from the state $A_{i-1}(t) = n_{i-1}-1$, $A_i(t) = n_i+1$ for some $i = 2, ..., K$; this is the state $L_i\boldsymbol{n}$. In this case, an agent jumping from compartment i into compartment $i-1$ is required for the system to arrive at state $\boldsymbol{n}$. The probability of this event occurring in an infinitesimal time step is equal to $d(n_i + 1)\mathrm{d}t$. We multiply this probability by $p(L_i\boldsymbol{n}, t)$, where $p(L_i\boldsymbol{n}, t)$ is the probability of the system being in state $L_i\boldsymbol{n}$ at time t. Likewise, the system could transition from state $R_i\boldsymbol{n}$ to $\boldsymbol{n}$ through an agent jumping out of compartment i into compartment $i+1$ for $i = 1, ..., K-1$, which occurs with a similar probability. Finally, it could be the case that the state $\boldsymbol{n}$ is maintained throughout the time step; an event which occurs with probability $(1 - 2dn_ip(\boldsymbol{n}, t))\mathrm{d}t$. The third and fourth sums in the CME represent the proliferative and degradative reactions respectively, and can be derived in much the same way as the first and second.

To derive an evolution equation for the mean of $A_i(t)$, we multiply the CME by n_i, then sum over all possible states $\boldsymbol{n}$. We can thus write the vector of means as

$$\boldsymbol{M}(t) = [M_1(t), ..., M_K(t)] = \sum_{\boldsymbol{n}} \boldsymbol{n}p(\boldsymbol{n}) \equiv \sum_{n_1=0}^{\infty} ... \sum_{n_K=0}^{\infty} \boldsymbol{n}p(\boldsymbol{n}). \tag{2.5}$$

Applying this technique to our example system, however, yilds a system in which the means of the $A_i(t)$s depend on the variance and covariances, resulting from the non-linearity of the degradation probabilities; the variances and covariances themselves then depend on higher-order moments. This is an intrinsic problem in attempting to describe mesoscopic behaviours deterministically; namely, that it is often impossible to derive a closed system of evolution equations. Nonetheless, this problem can be mitigated by making the (often poor) assumption that the variances and all higher moments are zero, known as the mean-field assumption. The degree to which the mean-field assumption is acceptable depends heavily upon the specific model and the chosen model parameters. There are many other methods for closing the system of equations, broadly known as moment closures, which can achieve a higher degree of accuracy over the rudimentary mean-field closure.

Nevertheless, for the purpose of simplicity, this derivation shall be built upon the mean-field assumption, ultimately resulting in the following system of evolution equa-

tions for the compartment means:

$$\frac{\partial M_1}{\partial t} = d\left(M_2 - M_1\right) + k_2 M_1 + k_1 h^{-1}\left(M_1 - M_1^2\right), \tag{2.6}$$

$$\frac{\partial M_i}{\partial t} = d\left(M_{i+1} - 2M_i + M_{i-1}\right) + k_2 M_i - k_1 h^{-1}\left(M_i - M_i^2\right), \tag{2.7}$$

$$\frac{\partial M_K}{\partial t} = d\left(M_{K-1} - M_K\right) + k_2 M_K - k_1 h^{-1}\left(M_K - M_K^2\right), \tag{2.8}$$

where $i = 2, ..., K-1$. As with the CME, deriving this full system of equations is laborious. Instead, to demonstrate the general technique, we will derive the term $d\left(M_{i+1} - 2M_i + M_{i-1}\right)$ in the evolution equation for M_j, $j = 2, ..., K-1$. We begin by multiplying the first two terms of the CME by n_j and summing over $\boldsymbol{n}$, giving

$$d\sum_{i=1}^{K-1}\sum_{\boldsymbol{n}}\left\{n_j(n_i+1)p(R_i\boldsymbol{n}) - n_j n_i p(\boldsymbol{n})\right\} + d\sum_{i=2}^{K}\sum_{\boldsymbol{n}}\left\{n_j(n_i+1)p(L_i\boldsymbol{n}) - n_j n_i p(\boldsymbol{n})\right\}. \tag{2.9}$$

The first sum can then be written as

$$\begin{aligned} d\sum_{i=1}^{K-1}\sum_{\boldsymbol{n}}\left\{n_j(n_i+1)p(R_i\boldsymbol{n}) - n_j n_i p(\boldsymbol{n})\right\} = {} & d\sum_{i=1}^{j-2}\sum_{\boldsymbol{n}}\left\{n_j(n_i+1)p(R_i\boldsymbol{n}) - n_j n_i p(\boldsymbol{n})\right\} \\ & + d\sum_{i=j+1}^{K-1}\sum_{\boldsymbol{n}}\left\{n_j(n_i+1)p(R_i\boldsymbol{n}) - n_j n_i p(\boldsymbol{n})\right\} \\ & + d\sum_{\boldsymbol{n}}\left\{n_j(n_{j-1}+1)p(R_{j-1}\boldsymbol{n}) - n_j n_{j-1} p(\boldsymbol{n})\right\} \\ & + d\sum_{\boldsymbol{n}}\left\{n_j(n_j+1)p(R_j\boldsymbol{n}) - n_j^2 p(\boldsymbol{n})\right\}. \end{aligned} \tag{2.10}$$

Noting that $\langle n_i\rangle = \sum_{\boldsymbol{n}} n_i p(\boldsymbol{n})$ and $\langle n_i n_j\rangle = \sum_{\boldsymbol{n}} n_i n_j p(\boldsymbol{n})$, it follows that

$$\sum_{n}\left\{n_j(n_i+1)p(R_i\boldsymbol{n}) - n_j n_i p(\boldsymbol{n})\right\} = 0 \tag{2.11}$$

for $j \neq i$, $j \neq i+1$, allowing us to rewrite (2.10) as

$$\begin{aligned} d\sum_{i=1}^{K-1}\sum_{\boldsymbol{n}}\left\{n_j(n_i+1)p(R_i\boldsymbol{n}) - n_j n_i p(\boldsymbol{n})\right\} = {} & d\sum_{\boldsymbol{n}}\left\{n_j(n_{j-1}+1)p(R_{j-1}\boldsymbol{n})\right\} - \langle n_j n_{j-1}\rangle \\ & + d\sum_{\boldsymbol{n}}\left\{n_j(n_j+1)p(R_j\boldsymbol{n})\right\} - \langle n_j^2\rangle. \end{aligned} \tag{2.12}$$

Finally, noticing that

$$n_j(n_{j-1}+1) = (n_j-1)(n_{j-1}+1) + (n_{j-1}+1), \tag{2.13}$$

and

$$n_j(n_j+1) = (n_j+1)^2 - (n_j+1), \tag{2.14}$$

we obtain

$$d\sum_{i=1}^{K-1}\sum_{\boldsymbol{n}}\{n_j(n_i+1)p(R_i\boldsymbol{n}) - n_jn_ip(\boldsymbol{n})\} = d(\langle n_{j-1}\rangle - \langle n_j\rangle) = d(M_{j-1}-M_j). \tag{2.15}$$

Proceeding in a similar fashion, it is established that

$$d\sum_{i=2}^{K}\sum_{\boldsymbol{n}}\{n_j(n_i+1)p(L_i\boldsymbol{n}) - n_jn_ip(\boldsymbol{n})\} = d(M_{j+1}-M_j). \tag{2.16}$$

Clearly, (2.15) and (2.16) sum to give $d(M_{j+1}-2M_j+M_{j-1})$, as required.

It is notable that this particular term represents the flux into compartment j resulting from diffusive jumps into (the term $d(M_{j+1}+M_{j+1})$) and out of (the term $d(-2M_j)$) the compartment. Diffusive reactions are uni-molecular, and therefore their corresponding terms do not contain any higher-order moments of n_j. This same technique of multiplying by n_j, summing over the states $\boldsymbol{n}$, and carefully rearranging the result can be applied to the other terms of the CME to construct the full system of mean-field ODEs; however, this results in higher-order moments for the bi-molecular reactions, necessitating the use of an appropriate moment closure to obtain a closed system of ODEs.

The corresponding deterministic description of this system is usually expressed in terms of concentration, which can be approximated by $hu(x_i,t) \approx M_i$, where x_i is the centre of compartment i. Dividing (2.7) through by h gives

$$\frac{\partial u}{\partial t}(x_i,t) \approx d(u(x_{i-1},t) - 2u(x_i,t) + u(x_{i+1},t)) + \alpha u(x_i,t)\left(1-\beta u(x_i,t)\right), \tag{2.17}$$

where $\alpha = k_2 + h^{-1}k_1$, and $\beta = (k_1h)/(k_2h+k_1)$. Finally, Taylor expanding the right-hand side of (2.17), and taking the joint limit as $h \to 0$, $t \to 0$, holding the quantity dh^2 constant, yields

$$\frac{\partial u}{\partial t} \approx dh^2\frac{\partial^2 u}{\partial x^2} + \alpha u(1-\beta u). \tag{2.18}$$

This general technique can be used to find the deterministic approximation of any compartment-based reaction network.

Algorithm 1. The Gillespie algorithm.

1. Generate two uniformly distributed random numbers r_1, r_2 from the interval (0,1).
2. For each reaction i, compute the propensity function $\alpha_i(t)$. Further, compute

$$\alpha_0 = \sum_{i=1}^{q} \alpha_i(t),$$

   where q denotes the number of possible events.
3. Calculate the time elapsed to the next reaction, given by

$$\tau = \frac{1}{\alpha_0} \ln\left(\frac{1}{r_1}\right)$$

4. Determine which reaction occurs at time $t + \tau$. In particular, find the index j such that

$$\frac{1}{\alpha_0} \sum_{i=1}^{j-1} \alpha_i(t) \leq r_2 < \frac{1}{\alpha_0} \sum_{i=1}^{j} \alpha_i(t),$$

   updating the corresponding agent counts accordingly. Finally, set $t = t + \tau$, and return to step (1).

2.1.2 Stochastic simulation algorithms

We now turn our attention to the simulation of a compartment-based model. There are many methods to achieve this; in this report, we will consider only the Gillespie direct method [16], however, it is worth mentioning some of the alternatives. Exact methods, named to reflect their ability to exactly determine solution trajectories of the CME, offer the most accuracy, but are often prohibitively expensive. These include methods such as the Gillespie direct method [16] and the next sub-volume method [8]. Inexact methods provide computational efficiency at the expense of, to some extent, accuracy. These include the τ-leaping method [17], in which a fixed time step τ is chosen, and the number of events occurring of each type within this time step is drawn from a Poisson distribution with appropriate mean. The R-leaping method [3] is similar; however, here a prescribed number of events is given, with the time step being drawn from an Erlang distribution.

The Gillespie direct method proceeds as follows. As previously mentioned, each event, whether it be a reaction or a diffusive jump, has associated with it some propen-

sity function $\alpha_j(t)$. If there are q possible events, we define

$$\alpha_0 = \sum_{i=1}^{q} \alpha_i(t) \tag{2.19}$$

to be the sum of all such propensity functions. An implicit assumption in the construction of the compartment-based method is that the time until each particular event is exponentially distributed. Therefore, the time τ until the occurrence of any event, is simply the minimum of these respective exponential waiting times, i.e. $\tau \sim \text{Exp}(\alpha_0)$. To sample from this distribution, we employ the inverse transform sampling method; that is, we take a realisation of the uniform random variable $r_1 \sim \text{U}(0,1)$ and calculate

$$\tau = \frac{1}{\alpha_0} \ln\left(\frac{1}{r_1}\right). \tag{2.20}$$

After calculating this time step, we must also determine which specific reaction has occurred. Each event occurs with probability proportional to its propensity function; therefore, we must find the index j such that

$$\frac{1}{\alpha_0} \sum_{i=1}^{j-1} \alpha_i(t) \leq r_2 < \frac{1}{\alpha_0} \sum_{i=1}^{j} \alpha_i(t).$$

This process continues for the desired amount of time, finally yielding a single realisation of the system.

2.2 Macroscopic modelling

Reaction-diffusion systems are typically modelled at the macroscopic scale through the use of either deterministic or stochastic partial differential equations. The theory of PDEs is expansive, and there exist myriad methods for numerically solving them, such as the finite-difference, finite-element, and finite-volume methods, which can be optimised for high levels of efficiency and accuracy. Further, many specialist software packages such as `MATLAB` and `ANSYS` possess in-built solvers for specific classes of PDE, making PDEs an accessible and viable modelling technique even in the absence of an in-depth knowledge of the underlying theory.

Despite the efficiency and relative simplicity of PDE models, they may fail to provide a full description of systems in which stochastic fluctuations contribute significantly to the dynamics. This is a common problem encountered in the modelling of small-scale reaction-diffusion systems. When the number of agents in the domain is large,

stochastic fluctuations can be considered to have a negligible effect; on the other hand, for small numbers of agents, the effect of randomness on their behaviour cannot be adequately captured by a deterministic PDE. A further issue present in PDE modelling, discussed in Section 2.1.1, arises when second- or higher-order reactions are present in the system. As discussed, the presence of these higher-order reactions mandates the use of an appropriate moment closure in order to derive a closed, finite system of mean-field ODEs from the CME of the mesoscopic model.

2.2.1 Deriving the reaction-diffusion equation

To demonstrate the use of PDEs for modelling reaction-diffusion systems, we derive the reaction-diffusion equation (1.1) on the domain $\Omega \subset \mathbb{R}$. Define the function $c(x,t)$ to be the density of agents at $x \in \Omega$ at time t. Further, define a function $f(c)$ which represents the net change in concentration due to any and all reactions that take place, and define $J(x,t)$ to be the net flux at $x \in \Omega$ at time t in the positive x direction. For an arbitrary interval $\Omega' = [a,b] \subset \Omega$, the change in the number of agents present in Ω' over an arbitrarily short time period $[t, t+\tau)$ can be expressed as

$$\begin{array}{ccccccc} \text{Number of} & & \text{Number of} & & \text{Flux into } \Omega' \text{ over} & & \text{Production of agents} \\ \text{agents in } \Omega' \text{ at} & = & \text{agents in } \Omega' & + & \text{the interval} & + & \text{within } \Omega' \text{ over the} \quad . \\ \text{time } t+\tau & & \text{at time } t & & [t, t+\tau) & & \text{interval } [t, t+\tau) \end{array}$$

This can be written in terms of the functions and variables already introduced, giving

$$\int_{\Omega'} c(x, t+\tau)\,\mathrm{d}x = \int_{\Omega'} c(x,t)\,\mathrm{d}t - (J(b,t) - J(a,t))\tau + \int_{\Omega'} f(c(x,t))\tau\,\mathrm{d}x. \tag{2.21}$$

Rearranging and dividing through by τ gives

$$\int_{\Omega'} \frac{c(x,t+\tau) - c(x,t)}{\tau}\,\mathrm{d}x = -(J(b,t) - J(a,t)) + \int_{\Omega'} f(c(x,t))\mathrm{d}x, \tag{2.22}$$

to which we apply the fundamental theorem of calculus, such that

$$\int_{\Omega'} \frac{c(x,t+\tau) - c(x,t)}{\tau}\,\mathrm{d}x = -\int_{\Omega'} \frac{\partial J}{\partial x}\,\mathrm{d}x + \int_{\Omega'} f(c(x,t))\,\mathrm{d}x. \tag{2.23}$$

Finally, taking the limit as $t \to 0$ and rearranging, we obtain

$$\int_{\Omega'} \left\{ \frac{\partial c}{\partial t} + \frac{\partial J}{\partial x} - f(x) \right\} \mathrm{d}x = 0. \tag{2.24}$$

Since equation (2.24) must hold for any arbitrary $\Omega' \subset \Omega$, it is concluded that the

integrand must be zero. This gives us the general PDE,

$$\frac{\partial c}{\partial t} = -\frac{\partial J}{\partial x} + f(c). \tag{2.25}$$

In the context of a reaction-diffusion system, the only flux present is that resulting from diffusion; to this end, we apply Fick's law of diffusion,

$$J(x,t) = -D\frac{\partial c}{\partial x}(x,t), \tag{2.26}$$

for diffusion coefficient D. Substitution of Fick's law into equation (2.25) yields the one-dimensional reaction-diffusion equation,

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2} + f(c). \tag{2.27}$$

2.2.2 The Crank-Nicolson finite difference method

Finite difference methods (FDMs) are a commonly used method for the solution of partial differential equations. These methods estimate the derivatives at a number of discrete grid points using finite difference approximations. To achieve this, derivatives are approximated via Taylor's theorem, which states that, for any function f with sufficiently well-behaved derivatives,

$$f(a+h) = f(a) + \frac{\mathrm{d}}{\mathrm{d}x}f(a)h + \mathcal{O}(h^2). \tag{2.28}$$

For small h, therefore, we can approximate the derivative of f by

$$\frac{\mathrm{d}f}{\mathrm{d}x}(a) \approx \frac{f(a+h) - f(a)}{h}. \tag{2.29}$$

In an analogous manner, one can derive finite difference approximations for higher-order derivatives; that is,

$$\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}(a) \approx \frac{f(a+h) - 2f(a) + f(a-h)}{h^2}. \tag{2.30}$$

A popular FDM for solving diffusion-like equations is the Crank-Nicolson method, proven to be unconditionally stable [31]. To demonstrate this method, suppose we wish to solve the diffusion equation

$$\frac{\partial u}{\partial t} = D\frac{\partial^2 u}{\partial x^2}, \tag{2.31}$$

on the domain $\Omega = [a, b]$, with zero-flux boundary conditions

$$\frac{\partial u}{\partial x}(a,t) = \frac{\partial u}{\partial x}(b,t) = 0, \quad \forall t \geq 0. \tag{2.32}$$

In doing so, the domain Ω is discretised into a set of points $a = x_1 < x_2 < ... < x_{N-1} < x_N = b$ with uniform spacing Δx; further, we discretise time into a sequence of equally spaced time steps $0 = t_0 < t_1 < ...$ with uniform spacing Δt. For brevity, we will write $u(x_i, t_j) = u_i^j$. The Crank-Nicolson (CN) method makes use of the following finite difference approximations

$$\frac{\partial u}{\partial t} \approx \frac{u_i^{j+1} - u_i^j}{\Delta t}, \tag{2.33}$$

$$\frac{\partial u}{\partial x} \approx \frac{(u_{i+1}^{j+1} - u_i^{j+1}) + (u_{i+1}^j - u_i^j)}{2\Delta x}, \tag{2.34}$$

$$\frac{\partial^2 u}{\partial x^2} \approx \frac{(u_{i+1}^{j+1} - 2u_i^{j+1} + u_{i-1}^{j+1}) + (u_{i+1}^j - 2u_i^j + u_{i-1}^j)}{2\Delta x^2}. \tag{2.35}$$

The CN discretisation for the diffusion equation is then

$$\frac{u_i^{j+1} - u_i^j}{\Delta t} = \frac{D}{2\Delta x^2}\left((u_{i+1}^{j+1} - 2u_i^{j+1} + u_{i-1}^{j+1}) + (u_{i+1}^j - 2u_i^j + u_{i-1}^j)\right). \tag{2.36}$$

Defining

$$r = \frac{D\Delta t}{2\Delta x^2}, \tag{2.37}$$

the discretisation (2.36) may be expressed in the more manageable form

$$-ru_{i+1}^{j+1} + (1+2r)u_i^{j+1} - ru_{i-1}^{j+1} = -ru_{i+1}^j + (1-2r)u_i^j - ru_{i-1}^j. \tag{2.38}$$

The CN method is an implicit method; that is, the solution at each time step depends on both the previous state and the new state. Implicit methods require the solution of an equation at each time step, and therefore incur a greater computational cost at each step as compared to explicit methods, in which the solution depends only upon the previous state. Implicit methods are advantageous, however, in that they typically require significantly fewer time steps in order to achieve the same accuracy as comparable explicit methods. It is noted from (2.38) that, in the case of the CN discretisation of the diffusion equation, the problem arising in the computation of the updated solution is tridiagonal, which can be solved far more efficiently than a general linear system.

One potential difficulty encountered in the construction of FDMs is the implementation of Neumann boundary conditions, such as those in (2.32); clearly, equation (2.38)

is meaningless when $i = 1, N$. The standard technique for handling Neumann boundary conditions within the FDM framework necessitates the introduction of 'ghost nodes' at $i = 0$ and $i = N+1$. In the diffusion equation example, we can discretise the boundary conditions (2.32) using a central difference approximation of the derivative,

$$\frac{\partial u}{\partial x}(a,t) = \frac{u_2^j - u_0^j}{\Delta x} = 0, \tag{2.39}$$

$$\frac{\partial u}{\partial x}(b,t) = \frac{u_{N+1}^j - u_{N-1}^j}{\Delta x} = 0, \tag{2.40}$$

which gives that

$$u_0^j = u_2^j, \tag{2.41}$$

$$u_{N+1}^j = u_{N-1}^j. \tag{2.42}$$

These can then be substituted into (2.38) to give the full finite difference scheme for the diffusion equation.

2.3 The pseudo-compartment method

In this section we present a detailed description of the pseudo-compartment method, on which we base the titular model of this report. The PCM computes approximate solutions of the general reaction-diffusion equation

$$\frac{\partial u}{\partial t} = D\nabla^2 u + R(u), \tag{2.43}$$

on some domain Ω. In this section, we take the domain $\Omega = [a, c] \subset \mathbb{R}$, noting that the extension of this method to higher dimensions is trivial. The domain is divided into two connected components, which we here define, without loss of generality, to be $\Omega_P = [a, b]$ and $\Omega_C = [b, c]$.

On the sub-domain Ω_P, we solve Equation (2.43). A zero-flux boundary condition is enforced at the interface b; this choice of boundary condition reflects the fact that the flux of agents between the two regions is not a deterministic process, but rather a discrete stochastic process. The appropriate boundary condition on a is chosen according to the specific application. The domain Ω_C is divided into compartments $C_1, ..., C_K$ of uniform width $h = (c - b)/K$ on which, as discussed in Section 2.1, any number of reactive or diffusive events can be defined. The PDE is solved according to the Crank-Nicolson finite difference method, with a grid spacing $\Delta x << h$ and time step $\mathrm{d}t_P$; we note, however, that any appropriate method can be used for solving this PDE,

as the algorithm for interfacing between the two sub-domains is independent of any specific implementation. Agents in Ω_C can jump between neighbouring compartments with propensity $A_i D/h^2$, where A_i is the number of agents in C_i, as well as undergo reactions according to any number of reaction channels with appropriate propensities; in Ω_P, these diffusive and reactive mechanisms are represented by the PDE itself.

The key mechanism for coupling the two sub-domains is the pseudo-compartment, defined to be the interval $C_{-1} = [b-h, b] \subset \Omega_P$. Agents in Ω_C are allowed to jump into the pseudo-compartment, with the usual propensity $A_i D/h^2$. Thus, the important consideration is how to correctly count the number of agents inside the pseudo-compartment; this is done by computing the expected value of A_{-1},

$$A_{-1} = \int_{C_{-1}} u(x,t)\mathrm{d}x.$$

Hence, when an agent jumps into C_{-1}, we uniformly add $1/h$ to the solution u at each PDE grid point in C_{-1}. Similarly, agents can jump out of the pseudo-compartment into C_1 with propensity $\alpha^* = A_{-1}D/h^2$, in which case we subtract an agent's worth of mass, i.e.

$$u(x, t+\tau) = u(x,t) - \mathbb{I}_{C_{-1}}/h. \tag{2.44}$$

The algorithm for the implementation of the PCM is similar to the standard Gillespie algorithm (Algorithm 1), with a slight modification made to concurrently simulate both the PDE and the compartments. The time is initialised by $t = 0$ and $t_P = t + \mathrm{d}t_P$. At each time step, the time of the next compartment-based reaction is calculated as $t_C = t + \tau$, with τ computed according to Equation (2.20). If $t_C < t_P$ then the compartment-based reaction occurs as in the Gillespie algorithm, and the time is updated to $t = t_C$; if, instead, $t_P < t_C$ then the PDE solution is updated via the Crank-Nicolson method, the current time is updated to $t = t_P$, and the next PDE update time is set to $t_P = t_P + \mathrm{d}t_P$. The algorithm then repeats from the new time t, ending once some pre-specified total time has elapsed.

The PCM has been demonstrated to be effective for representing a number of behaviours, such as maintaining a uniform steady state, and establishing a morphogen gradient across a domain [38]. Here, we present a novel application of the PCM; namely, the replication of a travelling wave solution to the Fisher equation on a finite domain. On Ω_C, we simulate the reaction system

$$A_1 \underset{d}{\overset{d}{\rightleftharpoons}} A_2 \underset{d}{\overset{d}{\rightleftharpoons}} \ldots \underset{d}{\overset{d}{\rightleftharpoons}} A_K, \tag{2.45}$$

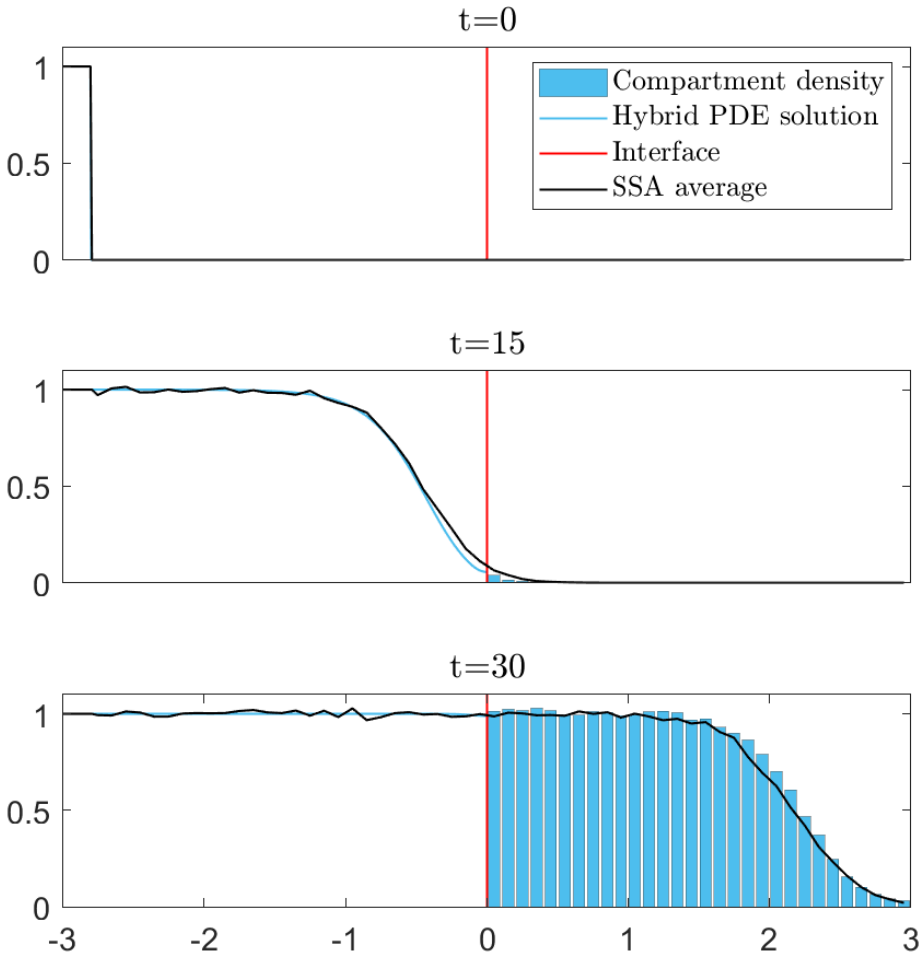

Figure 2-1: This plot shows the calculated density profile of the PCM at $t = 0, 15, 30$ (top, middle, and bottom resp). We take the domain $\Omega = [-3, 3]$ with regions $\Omega_P = [-3, 3]$, $\Omega_C = [0, 3]$. On Ω_P we numerically solve (2.47) using the Crank-Nicolson method with grid spacing $\Delta x = 0.005$ and time step $\Delta t = 0.0005$, as well as model parameters $D = 1/100$, $\lambda = 1$, and $u_\infty = 1000$. On Ω_C we simulate the compartment-based method with a compartment size of $h = 0.05$ and model parameters $k_1 = 1$, $k_2 = 0.01$, $d = 4$.

$$A_i + A_i \underset{k_2}{\overset{k_1}{\rightleftharpoons}} A_i, \tag{2.46}$$

where $d = D/h^2$, and k_1, k_2 are reaction rates. To derive the corresponding deterministic representation, we use the Poisson moment closure as in [25], which assumes that the mean of A_i is equal to its variance, as opposed to the mean-field closure, which

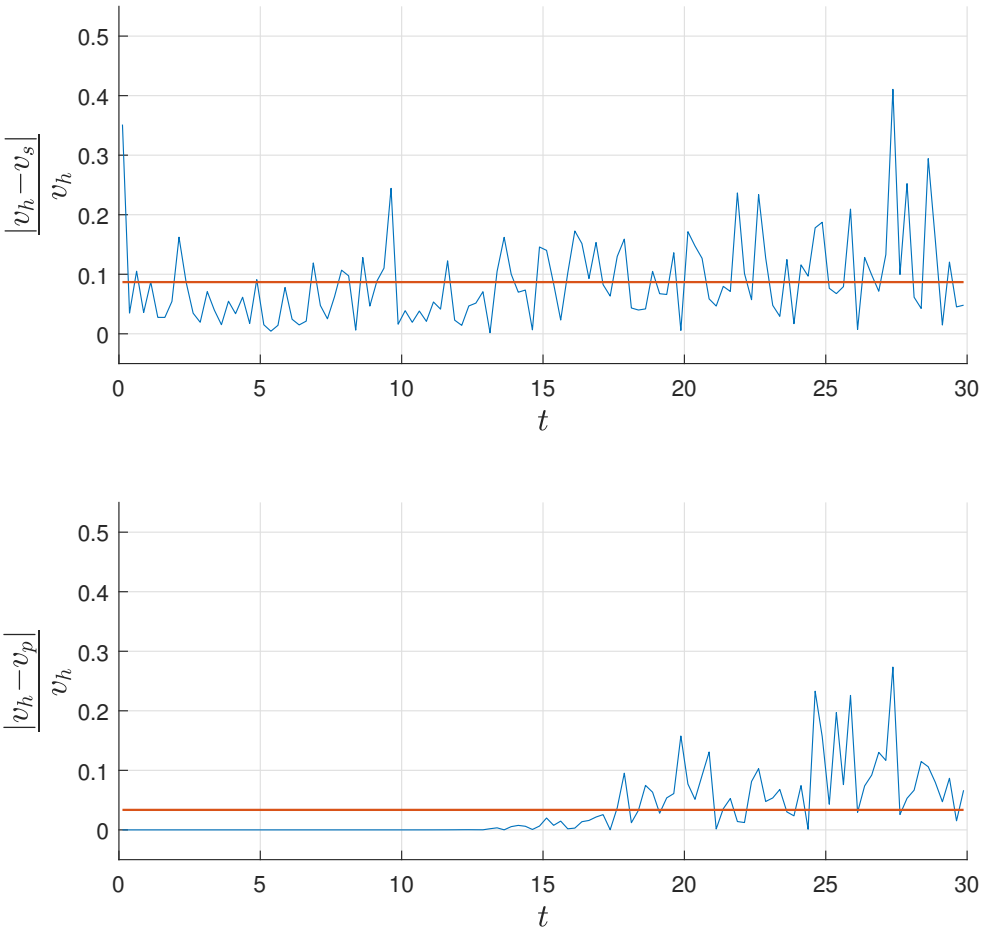

Figure 2-2: These plots show the absolute value of the relative error between the velocity of the travelling wave as predicted by the PCM and the SSA (top), and the PCM and the mean-field equations (bottom). Each point represents an average over 100 realisations of the PCM and 100 realisations of the SSA.

assumes that the variance is equal to zero. The result is the FKPP equation

$$\frac{\partial u}{\partial t} = D\frac{\partial^2 u}{\partial x^2} + \lambda u\left(1 - \frac{u}{u_\infty}\right), \tag{2.47}$$

which we solve on Ω_P with boundary conditions

$$\frac{\partial u}{\partial x}(a,t) = \frac{\partial u}{\partial x}(b,t) = 0. \tag{2.48}$$

In this case, the boundary conditions (2.48) are equivalent to the zero-flux boundary conditions. Note that this differs from the usual form of the FKPP equation, which is defined on an infinite domain.

To assess the effectiveness of the PCM in this instance, we compare the results obtained from the hybrid model against the fully stochastic simulation, which we consider to be the true solution of the problem. To begin, we plot the density profile as calculated by the PCM against the density profile as calculated by the SSA (Figure 2-1). We observe reasonable agreement between the two, with the PCM exhibiting a travelling wave of approximately the correct shape and velocity as that predicted by the SSA. While these preliminary results look promising, we acknowledge that a noticeable disagreement exists at the front of the travelling wave.

In Figure 2-2 we conduct a more quantitative evaluation of the agreement between the methods. Here, we compute the relative error between the velocity of the travelling wave as predicted by the PCM, v_h, and the predicted velocity of the SSA, v_s. We further compute the relative error between v_h and the velocity predicted by the mean-field equations, v_p. It is observed that the average errors are approximately 9% and 4%, respectively.

The purpose of the plots presented in Figure 2-2 is to assess whether there exists any systematic bias in the error between the two methods. In order to reasonably conclude that two methods predict the same behaviour, one would expect the relative errors between the two at each sample point to be identically distributed; clearly, the figure demonstrates that this is not the case. In particular, it would appear that in both comparisons the relative error increases with time. This is most evident when comparing the results of the PCM to the solution of the mean-field equations. Given the stationarity of the interface, it is expected that the relative error would initially be negligible, as stochastic effects do not manifest until agents begin to move into Ω_P at approximately $t = 15$. Another point of interest is the increase in the error as time continues past $t = 15$, which suggests that the PCM predicts a different wave velocity to the mean-field equations. This is also to be expected; as discussed previously, the presence of second-order reactions in our reaction network necessitates the use of a moment closure approximation, resulting in mean-field equations which are insufficient to satisfactorily describe the dynamics of the fully stochastic system.

The results of our comparison suggest that the PCM with a stationary interface is not suitable for modelling our reaction network; however, it is reasonable to assume that equipping the PCM with an adaptive interface could rectify many of these issues [5].

2.4 Multi-stage cell cycle model

As discussed, there are many ways in which the multi-stage cell cycle can be modelled. Here, we describe the deterministic, macroscopic model of Gavagnin *et al.* [13], which we will suitably adapt in Chapter 3. This model describes the volume-excluded diffusion and proliferation of a population of agents containing N stages on a $L_x \times L_y$ square lattice with grid spacing Δ and southwesterly lattice point located at the origin. At the individual level, agents are able to move from stage s to $s+1$ at rate λ_s; agents in the final stage N attempt proliferation events at a rate λ_N. To carry out a proliferation event, a random site is selected uniformly at random from the four nearest neighbours of the parent agent. When the chosen site is empty, a new agent in the first stage is placed in it, and the parent is set back to the first stage; when it is occupied, the event is aborted. Agent movement happens in a similar fashion; agents attempt movement events at rate β, choosing uniformly at random one of their neighbouring sites, only completing the event if the selected site is unoccupied.

To describe the overall population dynamics, we define $I_s(i,j,t) \in \{0,1\}$ to be the indicator that an agent in stage s occupies the lattice site $(i\Delta, j\Delta)$ at time t, where $i = 0, ..., L_x - 1$ and $j = 0, ..., L_y - 1$. It is noted that, in the case of periodic boundary conditions on the top and bottom of the lattice, as well as for initial conditions which are invariant under translations in the y direction, we can describe the overall dynamics in one dimension. To do this, we define the column density of s-stage agents $S_s(i,t)$ to be

$$S_s(i,t) = \frac{1}{L_y} \sum_{j=0}^{L_y-1} I_s(i,j,t). \tag{2.49}$$

We further define the total column density to be

$$C(i,t) = \sum_{s=1}^{N} S_s(i,t) = \frac{1}{L_y} \sum_{s=1}^{N} \sum_{j=0}^{L_y-1} I_s(i,j,t). \tag{2.50}$$

This continuum description of this agent-based model is derived by writing the master equation for the column density S_s and taking the limit as $\Delta \to 0$, where $\beta\Delta^2$ is held constant. To demonstrate the general technique for deriving this system of PDEs, we will consider the simplified case of a single agent stage, following a variation of the method of Simpson *et al.* [27, 26]. We define $\langle C_i(t) \rangle$ to be $C(i,t)$ averaged over a large number of repeats.

To begin, we write the discrete conservation equation for $\langle C_i(t) \rangle$, which represents the change in the average density of column i resulting from transitions into and out of column i over the interval $[t, t+\tau)$ for some infinitesimal time step τ. The probability of

a given agent experiencing a movement event in this time interval is $\beta\tau$. Likewise, the probability of a proliferation event is $\lambda\tau$. For brevity, we will here write $\langle C_i\rangle = \langle C_i(t)\rangle$; the conservation equation can then written as follows

$$\begin{aligned}\delta\langle C_i\rangle = &-\frac{\beta\tau}{4}\langle C_i\rangle(1-\langle C_{i-1}\rangle) - \frac{\beta\tau}{4}\langle C_i\rangle(1-\langle C_{i+1}\rangle)\\ &+\frac{\beta\tau}{4}\langle C_{i-1}\rangle(1-\langle C_i\rangle) + \frac{\beta\tau}{4}\langle C_{i+1}\rangle(1-\langle C_i\rangle)\\ &+\frac{\lambda\tau}{4}\langle C_{i-1}\rangle(1-\langle C_i\rangle) + \frac{\lambda\tau}{4}\langle C_{i+1}\rangle(1-\langle C_i\rangle)\\ &+\frac{\lambda\tau}{2}\langle C_i\rangle(1-\langle C_i\rangle),\end{aligned} \tag{2.51}$$

where $\delta\langle C_i\rangle$ denotes the change in average column density over an arbitrarily small time step of length τ. The first two terms correspond to diffusive jumps out of column i into columns $i-1$ and $i+1$, respectively. The second two terms represent diffusive jumps into column i from columns $i-1$ and $i+1$, respectively. Terms five and six represent the placement of daughter agents into column i from the neighbouring columns, and the final term represents the proliferation of agents strictly inside column i. Defining the average spatial column density to be $C(x_i,t) = \langle C_i\rangle$, where $x_i = \Delta i$, and simplifying (2.51) gives the more familiar form

$$\begin{aligned}C(x_i,t+\tau) - C(x_i,t) = &\frac{\beta\tau}{4}\left[C(x_{i-1},t) - 2C(x_i,t) + C(x_{i+1},t)\right]\\ &+\frac{\lambda\tau}{4}\left[C(x_{i-1},t) + 2C(x_i,t) + C(x_{i+1},t)\right]\left[1 - C(x_i,t)\right].\end{aligned} \tag{2.52}$$

Dividing through by τ and taking the Taylor expansion of both sides gives

$$\frac{\partial C}{\partial t} \approx \frac{\beta\Delta^2}{4}\frac{\partial^2 C}{\partial x^2} + \frac{\lambda}{4}\left[4C + \Delta^2\frac{\partial^2 C}{\partial x^2}\right][1-C] + \mathcal{O}(\Delta^2). \tag{2.53}$$

Finally, we jointly take the limit $\Delta \to 0$ and $\tau \to 0$, holding both $\beta\Delta^2$ and Δ^2/τ constant, to obtain

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} + \lambda(1-C)C. \tag{2.54}$$

The derivation of the continuum description in the case of an arbitrary number of stages proceeds in much the same way [26], resulting in the following system of

reaction-diffusion-type partial differential equations

$$\frac{\partial S_1}{\partial t} = D\frac{\partial}{\partial x}\left[(1-C)\frac{\partial S_1}{\partial x} + S_1\frac{\partial C}{\partial x}\right] + 2\lambda_N(1-C)S_N - \lambda_1 S_1, \tag{2.55}$$

$$\frac{\partial S_s}{\partial t} = D\frac{\partial}{\partial x}\left[(1-C)\frac{\partial S_s}{\partial x} + S_s\frac{\partial C}{\partial x}\right] + \lambda_{s-1}S_{s-1} - \lambda_s S_s, \tag{2.56}$$

$$\frac{\partial S_N}{\partial t} = D\frac{\partial}{\partial x}\left[(1-C)\frac{\partial S_N}{\partial x} + S_N\frac{\partial C}{\partial x}\right] + \lambda_{N-1}S_{N-1} - \lambda_N(1-C)S_N, \tag{2.57}$$

where $D = \lim_{\Delta\to 0}\frac{\beta\Delta^2}{4}$. The reflective boundary conditions enforced in the agent-based model then become zero-flux boundary conditions on each of the PDEs (2.55)-(2.57). Summing the above equations gives the dynamics of the total population,

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} + \lambda_N(1-C)S_N. \tag{2.58}$$

Note that, although the diffusive terms for each individual stage are non-linear, the population as a whole follows the same diffusion and growth rates as in the single-stage case.

Chapter 3

The hybrid method for multi-stage cellular invasion

In this chapter, we lay the foundations for building a hybrid method for cellular invasion with realistic cell cycle time distributions. This method will couple the agent-based model of Gavagnin *et al.* [13] with its corresponding continuum description, using a coupling method similar to that of Yates' [38] pseudo-compartment method. Further, we discuss a number of issues arising in the method, which require rectification prior to practical application.

3.1 The CCTD hybrid method

Let $\Omega_P = [a, b]$ denote the region in which we solve the system of S partial differential equations (2.55)-(2.57), with diffusion coefficient D and transition rates $\lambda_1, \lambda_2, ..., \lambda_S$. Moreover, we define Ω_L to be a $L_x \times L_y$ lattice with uniform grid spacing Δ upon which we simulate the agent-based model. The hybrid method is based on the PCM and, as such, there exists no deterministic flux between the two regions; for implementation, zero-flux boundary conditions are enforced on the PDE system.

Agents are permitted to traverse between the two regions only via the pseudo-compartment, which we define to be the sub-domain $C_{-1} = [b - \Delta, b]$. Note that here it is assumed, without loss of generality, that the region Ω_P lies to the left of Ω_L. To facilitate the coupling between the two regions, we must first determine the number of agents in each stage within the pseudo-compartment. Denote by $\mathcal{N}^s_{-1}(t)$ the number of agents at stage s in the pseudo-compartment, which is defined to be

$$\mathcal{N}^s_{-1}(t) = L_y \int_{C_{-1}} u_s(x, t)\, \mathrm{d}x, \tag{3.1}$$

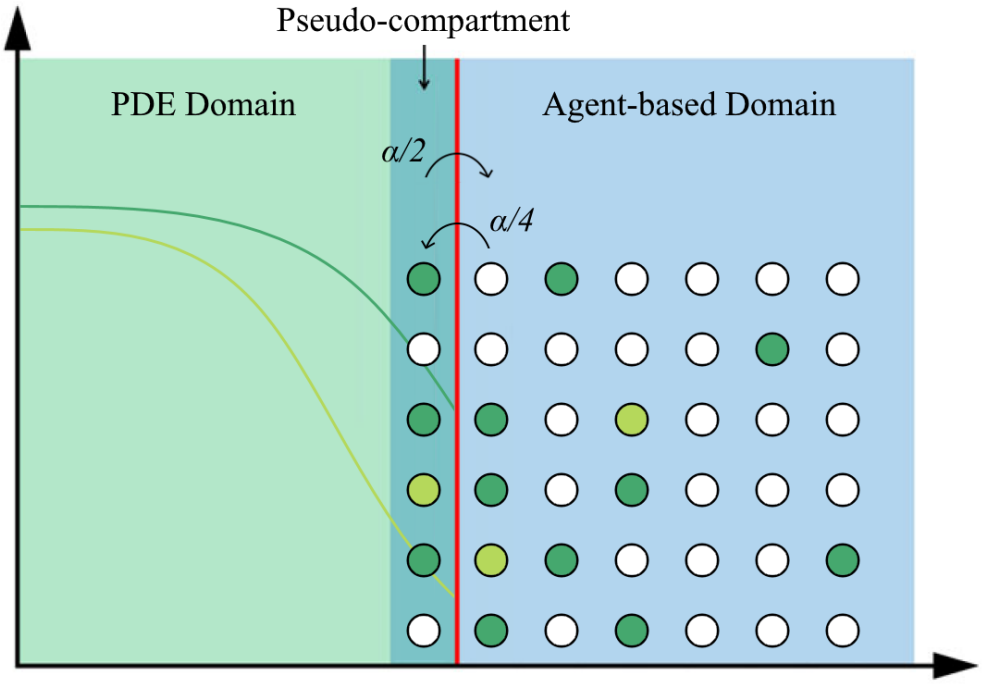

Figure 3-1: Schematic diagram for the multi-stage hybrid method. The green and yellow lines denote solutions to the system of PDEs (2.55)-(2.57) in the case where there are only two stages. The circles represent points on the lattice Ω_L. White circles are vacant, and the green and yellow circles represent agents in the first and second stages respectively. The arrows represent the propensities with which agents move across the interface, shown as the red line.

where the u_s are the solutions to the system (2.55)-(2.57).

Recall that agents on the lattice attempt movement events at rate $\beta = 4D/\Delta^2$; when a movement event is attempted, the associated movement direction is selected uniformly at random. If the neighbouring site in the selected direction is vacant, the agent moves into that site; otherwise, the event is aborted. If the moving agent lies on the right side of the lattice, any movement events to the right are reflected. If they are on the top of the lattice, any attempt to move upwards instead places them at the bottom of the lattice in the same column, imitating a periodic boundary condition. Likewise, downwards movements at the bottom of the lattice place the moving agent at the top. Agents move into the pseudo-compartment when they attempt to move leftwards off of the lattice; in this case, a single agent's worth of mass is added to the solution of the appropriate PDE in C_{-1}, i.e.

$$u_s(x, t+\tau) = u_s(x,t) + \mathbb{I}_{C_{-1}}/(\Delta L_y). \tag{3.2}$$

Algorithm 2. The multi-stage hybrid method.

1. Initialise the time $t = 0$, and set the final time t_f. Specify the PDE discretisation grid size Δx and time step $\mathrm{d}t_p$. Set the next PDE update time to be $t_p = t + \mathrm{d}t_p$. Set the number of stages S.

2. Initialise the mass distribution in the PDE region Ω_P and agents on the lattice Ω_L.

3. Calculate the propensity functions α_i, $i = 1, ..., S + 2$, according to equations (3.3)-(3.6).

4. Calculate the total propensity α_0 according to equation (3.7).

5. Determine the time until the next agent-based event $t_c = t + \tau$, where τ is calculated via equation (3.8).

6. If $t_c < t_p$
   (a) Determine which type of event occurs, where each type of event occurs with probability proportional to its propensity.
   (b) Select uniformly at random an agent, in the appropriate stage, that will attempt the event.
   (c) If the event is a transition event, advance the agent one stage. If the event is either a movement or proliferation, then attempt the event according to the mechanism given in the text.
   (d) Update the current time $t := t_c$.

7. Else if $t_p < t_c$
   (a) Update the PDE solutions according to the desired numerical method.
   (b) Update the current time $t := t_p$, and update the next PDE update time $t_p := t_p + \mathrm{d}t_p$.

8. If $t > t_f$, break. Otherwise, return to step (3).

Proliferation on the lattice occurs in a similar fashion. Agents in stage s, where $s < S$, progress to stage $s+1$ at a rate λ_s; agents at state S attempt proliferation events at rate λ_N. When an S-stage agent attempts to proliferate, it proceeds in much the same manner as if it were performing a movement event. Namely, a direction is chosen uniformly at random, and the occupancy of the neighbouring site is assessed. If the site is occupied, the event is aborted. If, however, the site is vacant, a new agent in the first stage is placed on that site, and the original agent is sent back to the first stage. Boundary conditions are the same as for movement on the top, right, and bottom sides of the lattice; however, agents are not permitted to proliferate leftwards into the pseudo-compartment, and hence a reflective boundary is imposed on the left side.

Agents can enter Ω_L from the pseudo-compartment via diffusion. Agents in stage s attempt to move into the leftmost column of Ω_L at a rate governed by the propensity function

$$\alpha_1 = \frac{1}{2}\mathcal{N}^s_{-1}(t). \tag{3.3}$$

Notice the factor of $1/2$ here, which represents a reflective boundary condition on agents attempting to jump into Ω_L. When such an event is attempted, a site in the leftmost column of Ω_L is selected uniformly at random. If the site is occupied, the event is abandoned. If the site is vacant, an agent at stage s is placed on the selected site, and an agent's worth of mass is removed from the solution u_s in C_{-1}, i.e.

$$u_s(x, t+\tau) = u_s(x,t) - \frac{1}{\Delta L_y}\mathbb{I}_{C_{-1}}. \tag{3.4}$$

Time proceeds in the hybrid method in much the same way as in the PCM. Define $\mathcal{N}^s(t)$ to be the number of s-stage agents in Ω_C at time t, and $\mathcal{N}(t) = \sum_s \mathcal{N}^s(t)$. Since all agents possess the same movement rate, we can assign a single propensity function to the event that a movement occurs, given by

$$\alpha_2 = \beta\mathcal{N}(t). \tag{3.5}$$

Similarly, all agents within a particular stage share the same transition rate. Therefore, with each type of transition event (including the proliferation event), we can associate the propensity

$$\alpha_{2+s} = \lambda_s\mathcal{N}^s(t), \tag{3.6}$$

for $s = 1, ..., S$. In order to calculate the time until the next event of any type occurs,

we first compute the sum of the propensity functions

$$\alpha_0 = \sum_{i=1}^{S+2} \alpha_i, \tag{3.7}$$

and calculate the time step

$$\tau = \frac{1}{\alpha_0} \ln \frac{1}{r}, \tag{3.8}$$

where $r \sim U(0, 1)$.

The full algorithm for the hybrid method is given in Algorithm 2, and a schematic diagram can be found in Figure 3-1.

3.2 Results

To determine the viability of the hybrid method, we conduct a simple experiment to evaluate the predictive accuracy of the method, with respect to the expected behaviour as determined by the fully stochastic agent-based model. In particular, we restrict ourselves to the case of a single stage, with a proliferation rate of $\lambda = 0$. This case corresponds precisely to the equation

$$\frac{\partial S}{\partial t} = D \frac{\partial^2 S}{\partial x^2}, \tag{3.9}$$

with zero flux boundary conditions. In Figure 3-2, we compare the density of the hybrid solution to that predicted by the mean-field equations. While, at a glance, the agreement seems good, there is a noticeable discontinuity across the interface. The result of this is emphasised in Figure 3-3, where we plot the expected total mass in Ω_P according to the mean-field solution, in contrast to that calculated by the hybrid method. Inspection reveals that the rate at which mass leaves Ω_P appears to be the same in both cases after approximately $t = 20$. One proposed explanation for this is that the coupling mechanism is unable to appropriately manage the discontinuity present in the initial condition, resulting in a lower flux into Ω_L than predicted.

The hybrid method, in its current state, is unable to simulate the full multi-stage system. The method suffers from a number of problems, both in terms of speed and numerical accuracy. It is hypothesised that these issues stem from inherent problems in the chosen PDE solver; the `pdepe` solver is unable to satisfactorily compensate for the discontinuities incurred by the removal of agents from the pseudo-compartment, resulting in solutions which violate the necessary zero-flux boundary conditions. In particular, the mass in Ω_P is not conserved between time steps. Further, numerical

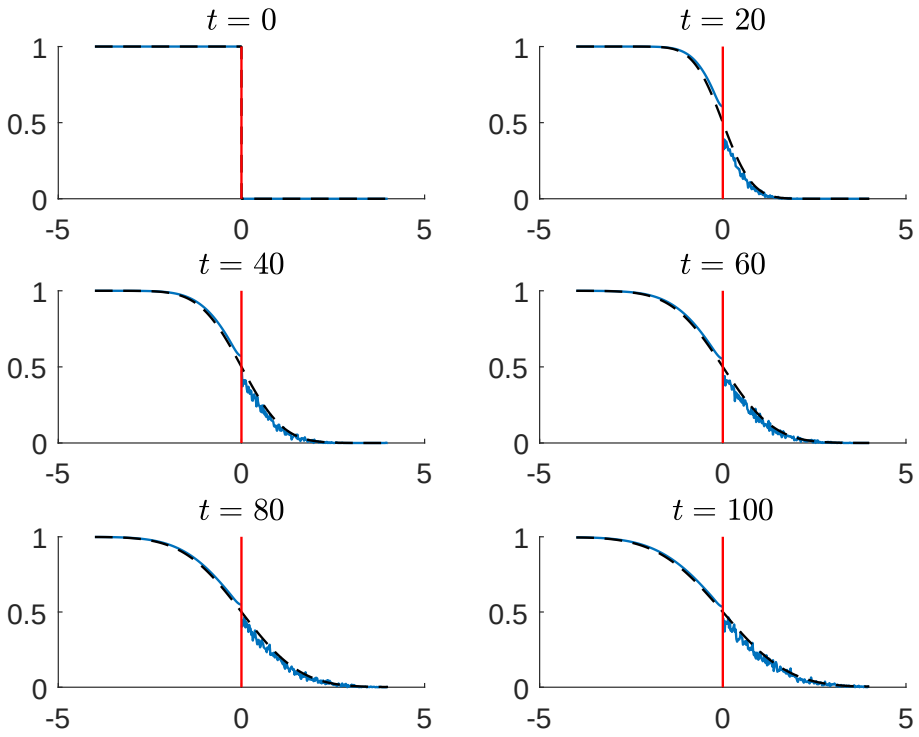

Figure 3-2: Mass is initially distributed entirely in Ω_P and permitted to diffuse throughout the domain with diffusion coefficient $D = 1/100$. The red line represents the location of the interface; to the left is Ω_P, and to the right, Ω_L. In Ω_L, we plot the average column density over 100 repetitions on a lattice of size 80 by 400. On Ω_P, we solve (3.9) with zero-flux boundary conditions using `MATLAB`'s built-in PDE solver `pdepe`.

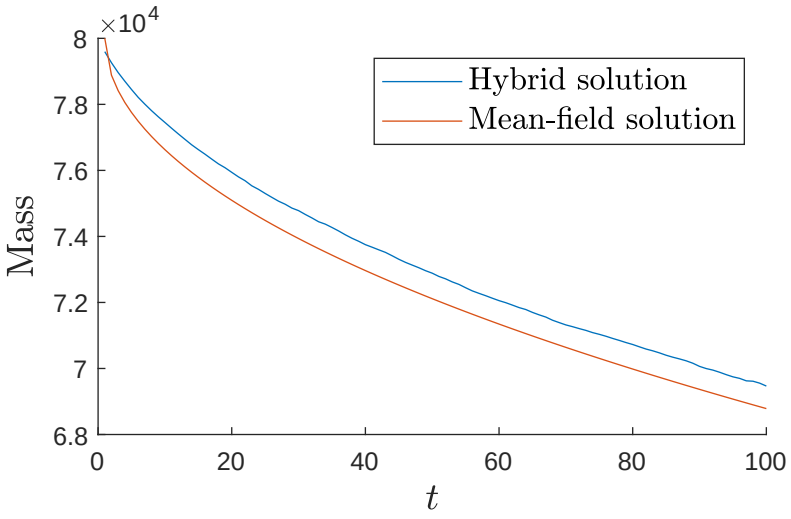

Figure 3-3: Here we plot the error for the diffusion problem, in which all agents are initially located in Ω_P. The blue line corresponds to the average mass in Ω_P as calculated by the hybrid method, averaged over 100 repetitions. The orange line corresponds to the expected value of the mass in Ω_P according to the mean-field model.

testing conducted towards the present work has revealed that the use of an explicit finite-difference method for solving the system of PDEs requires prohibitively small time steps in order to achieve stability.

It is also reasonable to conjecture that the location of the interface plays an important role in ensuring that the correct flux between the two regions is maintained. Upon inspection of Figure 3-3, it appears that a steep gradient at the interface may result in a dampening of the flux from the PDE-based region into the agent-based region; further numerical experiments are required to investigate this phenomenon.

Chapter 4

Discussion and conclusion

In this chapter, we present a brief discussion of the results and observed performance of the multi-stage hybrid method. Following this, we consider potential research avenues towards improvements in both the accuracy and generalisability of the method, alongside prospective projects to be undertaken during the PhD.

4.1 Discussion

In Chapter 3, we proposed an algorithm for simulating a multi-stage model of the cell cycle with realistic cell cycle time distributions, proliferation, and diffusion. This algorithm couples a system of reaction-diffusion type partial differential equations in one region with an on-lattice agent-based model in another. The method for interfacing between the two regions possesses great similarity to that used in the pseudo-compartment method [38].

We have demonstrated the potential of the method for coupling the two regimes; however, it is apparent that a significant amount of further development is necessary for the method to be considered viable. The potential improvements will be discussed further in Section 4.2, although a brief overview is given here for completeness. The first proposed improvement is to employ an implicit finite-difference scheme for solving the system of PDEs, with the hope of combating the observed numerical instability resulting from the use of an explicit scheme. The second is to replace the on-lattice method used in the mesoscopic regime with a volume-excluding compartment-based method [30]. Finally, the use of an adaptive interface could be considered, so as to ensure that the PDE system is not being used for the modelling of regions where agent numbers are low.

4.2 Future avenues of research

We will now discuss some of the avenues which could be explored in the PhD. On the whole, the PhD will be focused on hybrid methods for simulating reaction-diffusion systems. Particular focus will be given to developing novel techniques for simulating complex biological behaviours, such as cellular invasion and the cell-division cycle. The project will also emphasise novel applications of hybrid methods to real biological data sets; something which has not received a significant amount of attention in the past.

To date, there has been no attempt in the literature to couple a multi-stage system of PDEs with a mesoscopic method in a hybrid method. Therefore, the immediate aim of the PhD, following this report, is the improvement of our current hybrid methodology; as seen in Chapter 3, the method shows promise for the effective simulation of multi-stage diffusion systems. Further, the agent-based model and PDE-based model are both capable, independently, of demonstrating travelling wave behaviour [13]. The development of a hybrid method which is capable of simulating both multi-stage representations of the cell cycle as well as invasive behaviours would be both a valuable and novel addition to the current body of research, possessing applicability to a wide variety of biological systems such as wound healing [6], tumour formation [19], and embryonic development [15, 22].

There are improvements to be made with regards to the numerical methods for solving systems of PDEs representing multi-stage models of the cell cycle. While an explicit finite difference scheme suffices for solving the PDEs in isolation [13], it is evident from the results of Chapter 3 that this approach is not necessarily well-suited to coupling with a mesoscopic representation. It is unclear at present why this is the case, which leaves many open questions. One approach to combat the numerical instabilities caused by the current coupling method is the use of an implicit finite difference scheme, such as the Crank-Nicolson method, for solving the system of PDEs. An implicit finite difference scheme does not require the prohibitively small time steps necessary to ensure stability that an explicit method does. There are, however, other considerations to be made. For example, spurious solutions are still possible in the Crank-Nicolson method in the case that the ratio of $D\Delta t$ to Δx^2 exceeds $1/2$, as per a Von Neumann stability analysis [24]. The use of a spatio-temporal finite element method could also be considered [21]. The nonlinearity of the system of PDEs poses a significant computational strain on any numerical method; finding a method which is both efficient and well-behaved when interfacing with a mesoscopic model is of much interest.

The hybrid method constructed in this report couples a system of PDEs to an on-

lattice agent-based model. The justification for using the on-lattice method is that it provides a simple means to simulate volume exclusion; however, there are other ways to simulate this. For example, we could replace the on-lattice method with a volume-excluded compartment-based method. A common approach to achieve volume exclusion in a compartment-based method is to define the jump rates between compartments as the product of the diffusive jump rate with some blocking probability [30], for example, defining

$$d_i^{\pm} = \frac{D}{m^2 h^2}\left(1 - \frac{A_{i\pm1}}{m}\right), \tag{4.1}$$

where $d_i^{\pm}$ is the right (respectively left) volume-excluding jump rate of agents in compartment i and m is the carrying capacity of the compartments. Note that in the case $m = 1$, this corresponds exactly to a one-dimensional fully-excluding on-lattice method. The case where $m > 1$ is known as partial volume exclusion. Given the many different methods for incorporating volume exclusion into a mesoscopic model, there is much potential for developing efficient and accurate multi-stage models of reaction-diffusion systems, which the PhD will investigate.

Adaptive interfaces have seen wide implementation in many different hybrid methods, and their use has been demonstrated to substantially increase computational efficiency and accuracy for a number of problems [25, 29]. As demonstrated in our results from both the PCM and the multi-stage hybrid method, adaptive interfaces have the potential to solve the issues resulting from a coupling that is unable to respond to local changes in agent density.

4.3 Conclusion

A significant amount of work has been put into the development of hybrid methods; however, their widespread adoption has yet to be seen within the field of mathematical biology. There are several reasons for this. It is not immediately clear that hybrid methods are able to provide valuable insights into biological behaviours that traditional methods are unable to realise. Further, there exists little literature on the application of hybrid methods to growing domains, which are of great importance in developmental biology. The observed issues with the hybrid method constructed in this report signpost a wide variety of research directions, investigation into which will significantly benefit the field of hybrid modelling on the whole.

Bibliography

[1] Alberts, B., Bray, D., Hopkin, K., Johnson, A. D., Lewis, J., Raff, M., Roberts, K., and Walter, P. *Essential cell biology.* Garland Science, 2013.

[2] Anderson, A. R., and Chaplain, M. Continuous and discrete mathematical models of tumor-induced angiogenesis. *B. Math. Biol. 60*, 5 (1998), 857–899.

[3] Auger, A., Chatelain, P., and Koumoutsakos, P. R-leaping: Accelerating the stochastic simulation algorithm by reaction leaps. *J. Phys. Chem. A 125*, 8 (2006), 084103.

[4] Chao, H. X., Fakhreddin, R. I., Shimerov, H. K., Kumar, R. J., Gupta, G. P., and Purvis, J. E. Evidence that the cell cycle is a series of uncoupled, memoryless phases. *Preprint biorXiv, doi:10.1101/283614* (2018).

[5] de la Cruz, R., Guerrero, P., Calvo, J., and Alarcón, T. Coarse-graining and hybrid methods for efficient simulation of stochastic multi-scale models of tumour growth. *J. Comput. Phys. 350* (2017), 974–991.

[6] Deng, M., Chen, W.-L., Takatori, A., Peng, Z., Zhang, L., Mongan, M., Parthasarathy, R., Sartor, M., Miller, M., Yang, J., et al. A role for the mitogen-activated protein kinase kinase kinase 1 in epithelial wound healing. *Mol. Biol. Cell 17*, 8 (2006), 3446–3455.

[7] Deutsch, A., Dormann, S., and Maini, P. *Cellular automaton modeling of biological pattern formation: Characterization, applications, and analysis. 2005.* Birkhäuser, Boston (Basel).

[8] Elf, J., and Ehrenberg, M. Spontaneous separation of bi-stable biochemical systems into spatial domains of opposite phases. *Systems Biol. 1*, 2 (2004), 230–236.

[9] Elliott, E. C., and Cornell, S. J. Dispersal polymorphism and the speed of biological invasions. *PLoS One. 7*, 7 (2012), e40496.

[10] Erban, R., and Chapman, S. J. Stochastic modelling of reaction–diffusion processes: algorithms for bimolecular reactions. *Phys. Biol. 6*, 4 (2009), 046001.

[11] Fick, A. Ueber diffusion. *Ann. Phys. (Berl.) 170*, 1 (1855), 59–86.

[12] Fisher, R. A. The wave of advance of advantageous genes. *Ann. Eugenic. 7*, 4 (1937), 355–369.

[13] Gavagnin, E., Ford, M. J., Mort, R. L., T., R., and Yates, C. A. The invasion speed of cell migration models with realistic cell cycle time distributions. *J. Theor. Biol.* (2018).

[14] Gibson, M. A., and Bruck, J. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem. A 104*, 9 (2000), 1876–1889.

[15] Gilbert, S. F. The morphogenesis of evolutionary developmental biology. *Int. J. Dev. Biol. 47*, 7-8 (2003), 467.

[16] Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem. A 81*, 25 (1977), 2340–2361.

[17] Gillespie, D. T. A rigorous derivation of the chemical master equation. *Physica A 188*, 1-3 (1992), 404–425.

[18] Golubev, A. Applications and implications of the exponentially modified gamma distribution as a model for time variabilities related to cell proliferation and gene expression. *J. Theor. Biol. 393* (2016), 203–217.

[19] Hanahan, D., and Weinberg, R. A. The hallmarks of cancer. *Cell 100*, 1 (2000), 57–70.

[20] Isaacson, S. A. A convergent reaction-diffusion master equation. *J. Chem. Phys. 139*, 5 (2013), 054101.

[21] John, V., and Schmeyer, E. Finite element methods for time-dependent convection–diffusion–reaction equations with small diffusion. *Comput. Method. Appl. M. 198*, 3-4 (2008), 475–494.

[22] Keller, R. Cell migration during gastrulation. *Curr. Opin. Cell Biol. 17*, 5 (2005), 533–541.

[23] McQuarrie, D. A. Stochastic approach to chemical kinetics. *J. App. Probab. 4*, 3 (1967), 413–478.

[24] Oishi, C. M., Yuan, J. Y., Cuminato, J. A., and Stewart, D. E. Stability analysis of crank–nicolson and euler schemes for time-dependent diffusion equations. *BIT 55*, 2 (2015), 487–513.

[25] Robinson, M., Flegg, M., and Erban, R. Adaptive two-regime method: application to front propagation. *J. Chem. Phys. 140*, 12 (2014), 124109.

[26] Simpson, M., Landman, K., and Hughes, B. Multi-species simple exclusion processes. *Physica A 388*, 4 (2009), 399–406.

[27] Simpson, M., Landman, K., and Hughes, B. Cell invasion with proliferation mechanisms motivated by time-lapse data. *Physica A 389*, 18 (2010), 3779–3790.

[28] Smith, C. A., and Yates, C. A. Spatially extended hybrid methods: a review. *J. Royal Soc. Interface 15*, 139 (2018), 20170931.

[29] Spill, F., Guerrero, P., Alarcon, T., Maini, P. K., and Byrne, H. Hybrid approaches for multiple-species stochastic reaction–diffusion models. *J. Comput. Phys. 299* (2015), 429–445.

[30] Taylor, P. R., Baker, R. E., Simpson, M. J., and Yates, C. A. Coupling volume-excluding compartment-based models of diffusion at different scales: Voronoi and pseudo-compartment approaches. *J. Roy. Soc. Interface 13*, 120 (2016), 20160336.

[31] Thomas, J. W. *Numerical partial differential equations: finite difference methods*, vol. 22. Springer Science & Business Media, 2013.

[32] Tikhomirov, V. A study of the diffusion equation with increase in the amount of substance, and its application to a biological problem. In *Selected works of AN Kolmogorov*. Springer, 1991, pp. 242–270.

[33] van Zon, J. S., and Ten Wolde, P. R. Greens-function reaction dynamics: a particle-based approach for simulating biochemical networks in time and space. *J. Chem. Phys. 123*, 23 (2005), 234910.

[34] Von Smoluchowski, M. Mathematical theory of the kinetics of the coagulation of colloidal solutions. *J. Phys. Chem 92* (1917), 129–168.

[35] Weber, T. S., Jaehnert, I., Schichor, C., Or-Guil, M., and Carneiro, J. Quantifying the length and variance of the eukaryotic cell cycle phases by a stochastic model and dual nucleoside pulse labelling. *PLoS Comput. Biol. 10*, 7 (2014), e1003616.

[36] Wise, S. M., Lowengrub, J. S., Frieboes, H. B., and Cristini, V. Three-dimensional multispecies nonlinear tumor growthi: model and numerical method. *J. Theor. Biol. 253*, 3 (2008), 524–543.

[37] Woolley, T. E., Baker, R. E., and Maini, P. K. Turings theory of morphogenesis: where we started, where we are and where we want to go. 219–235.

[38] Yates, C. A., and Flegg, M. B. The pseudo-compartment method for coupling partial differential equation and compartment-based models of diffusion. *J. Royal Soc. Interface 12*, 106 (2015), 20150141.

[39] Yates, C. A., Ford, M. J., and Mort, R. L. A multi-stage representation of cell proliferation as a markov process. *B. Math. Biol. 79*, 12 (2017), 2905–2928.